Article:
Silva, M, Justeau, P, Rodrigues, S et al. (12 more authors) (2019) Untangling Neolithic and Bronze Age mitochondrial lineages in South Asia. Annals of Human Biology. pp. 1-5. ISSN 1464-5033

https://doi.org/10.1080/03014460.2019.1623319

Untangling Neolithic and Bronze Age mitochondrial lineages in South Asia

Silva, M.[a], Justeau, P.[a], Rodrigues, S.[a], Oteo-Garcia, G.[a], Dulias, K.[a], Foody, G.[a], Fichera, A.[a], Yau, B.[a], Rito, T.[b,c], Wilson, J.F.[d,e], Gandini, F.[a], Edwards, C.J.[a], Pala, M.[a], Soares, P.A.[f,g] and Richards, M.B.[a]*

[a]*Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, United Kingdom;* [b]*Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, 4710-057 Braga, Portugal;*[c]*ICVS/3B's, PT Government Associate Laboratory, 4710-057 Braga/4806-909 Guimarães, Portugal;* [d]*Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland, UK;* [e]*MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland, UK;* [f]*CBMA (Centre of Molecular and Environmental Biology), Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal;* [g] *Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*

*corresponding author: M.B.Richards@hud.ac.uk

Two key moments shaped the extant South Asian gene pool within the last 10 thousand years (ka): the Neolithic period, with the advent of agriculture and the rise of the Harappan/Indus Valley Civilisation, and Late Bronze Age events that witnessed the abrupt fall of the Harappan Civilisation and the arrival of Indo-European speakers. Here we focus on the phylogeographic patterns of mitochondrial haplogroups H2 and H13 in the Indian Subcontinent and incorporate evidence from recently released ancient genomes from Central and South Asia. We found signals of Neolithic arrivals from Iran, and later movements in the Bronze Age from Central Asia that derived ultimately from the Steppe. This study shows how a detailed mtDNA phylogeographic approach, combining both modern and ancient variation, can provide evidence of population movements, even in a scenario of strong male-bias such as in the case of the Bronze Age Steppe dispersals.

Keywords: mtDNA; Neolithic; Steppe; Bronze Age; South Asia; Indo-European

Introduction

We recently proposed a chronology for the peopling of South Asia by combining information from uniparental markers and genome-wide patterns (Silva et al. 2017). After the initial settlement in the Pleistocene, and the dispersal of some lineages south-eastwards from refugial areas in the Near East after the Last Glacial Maximum, two key moments brought additional layers of diversity to South Asia: (i) the Neolithic period, with the advent of farming and the rise of the Harappan/Indus Valley Civilisation in the north-west of the Subcontinent; and (ii) the upheavals in the Late Bronze Age that followed the 4.2 ka BP climatic event (which caused a dramatic increase in aridity in the Indus Valley region) (Staubwasser et al. 2003), and comprised the abrupt end of the Harappan Civilisation and the probable arrival of the Indo-European languages to the region (Parpola 2015).

Here, we take advantage of published ancient mitogenomes (whole mitochondrial DNA, or mtDNA, genomes) from Eurasia (Haak et al. 2015; Mathieson et al. 2015; Olalde et al. 2018), with special focus on the recently reported samples from Central and South Asia (Narasimhan et al. 2018), and reassess the findings advanced in our previous work (Silva et al. 2017) by focusing on two distinct lineages, whose distributions exceptionally span Europe, the Near East and South Asia: H2 and H13. Haplogroup H is one of the most common lineages found in present-day European populations (~45%) (Soares et al. 2010; Pala et al. 2016), but it is also found in the Near East, Caucasus, Central and South Asia, Siberia and North Africa (Loogväli et al. 2004). Given its presence at considerable frequencies in the European genetic substrate since the Neolithic (Brandt et al. 2013), it was most likely involved in diverse population movements in Eurasia, and can provide evidence of demographic events connecting Europe and Central/South Asia.

Methods

We built phylogenetic trees for each haplogroup using MtPhyl v5.003 (http://eltsov.org), guided by PhyloTree Build 17 (van Oven and Kayser 2009). We calculated Maximum Likelihood (ML) node age estimates with baseml v.4.7 from PAML package (Yang 1997), using a mutation rate of one substitution in every 3,624 years, and correcting for purifying selection (Soares et al. 2009). We used a total dataset of 621 (H2) and 326 (H13) modern sequences, and applied the HKY85 (Hasegawa, Kishino and Yano, 1995) mutation model with gamma-distributed rates (discrete distribution of 32 categories), considering two partitions to differentiate HVS-I and HVS-II from

the rest of the molecule. The age estimates reported in the Results and Discussion section for specific H2 and H13 branches result from this analysis.

Results and Discussion

With a Late Glacial age estimate of ~14 ka (95% confidence interval [0.9–19.3] ka), H2 most likely originated in Eastern Europe or the Caucasus and is divided into three main branches: H2a (~11 [7.9-–13.6] ka), H2b and H2c (both ~8 ka [4.9–10.2] and [2.0–14.5] ka, respectively). The oldest H2 genome known is a H2a1 lineage from the Russian Steppe Eneolithic (Mathieson et al. 2018), dating to earlier than 6 ka, with other H2a1 lineages from the Armenian Chalcolithic and Ukrainian Eneolithic dating to ~6 ka (Lazaridis et al. 2016; Mathieson et al. 2018). A basal H2 was also found in the Russian Steppe Eneolithic (Wang et al. 2018). H2a is by far the largest and most complex branch, comprising ~95% of the modern H2 mitogenome sequences in our dataset. After 5 ka, H2a appears in the Corded Ware in Poland (Juras et al. 2018) and in the Bell Beaker culture of Germany, reaching western Europe with the Copper or Bronze Age at ~4 ka (Allentoft et al. 2015). Additional sequences of H2a have been retrieved from Bronze and Iron Age individuals from the Pontic-Caspian steppe (Krzewińska et al. 2018). This earlier H2/H2a seems to be distributed around the Black Sea, before being dispersed from the Steppe with the Yamnaya pastoralist expansions across northern Europe.

H2b, on the other hand, is a minor branch (Figure 1). It contains several ancient samples from Russia, all basal to the rest of the branch, including one individual from the Yamnaya culture (I0441: 3010-2622 cal. BCE) (Haak et al. 2015) and one from the Late Bronze Age Srubnaya culture (I0431: 1850-1600 BCE) (Mathieson et al. 2015), both from the Pontic-Caspian Steppe region, and five other Bronze Age samples from east of the Volga river: three from Sintashta (I1024, I0942: 2050-1650 BCE, and I1053: 1922-1763 cal. BCE) and two from Krasnoyarsk (I1851: 1611-1459 cal. BCE and I1853: 1611-1503 cal. BCE) (Narasimhan et al. 2018). Also in a basal position, there are three modern Russian samples (two from the Altai region) and one Danish sequence (Lippold et al. 2014; Derenko et al. 2014; Li et al. 2014). Interestingly, while the vast majority (~70%) of H2 modern sequences in our dataset are of European origin, H2b displays a strong South Asian component, with seven samples from Pakistan, India and Sri Lanka. The newly published Sintashta and Middle Bronze Age Krasnoyarsk (Russian) sequences (Narasimhan et al. 2018), together with the previously released Yamnaya and Srubnaya, span a period from ~5 to 3.5 ka. These, plus the modern South Asian sequences, support our earlier suggestion that H2b was involved in movements east and southwards from the Pontic-Caspian region into South Asia (Silva et al. 2017), by documenting its progress eastwards across the

Eurasian Steppe. The Sintashta Culture in the Ural Mountains, or a "Sintashta-derived" culture (such as the Andronovo), is thought to have expanded eastwards into Central Asia ~3.8 ka, reaching South Asia within several hundred years (Gimbutas 1963; Anthony et al. 1986) and, based on linguistic and archaeological evidence, they are thought to have been responsible for spreading the Indo-European language family across Central and South Asia (Parpola 2015).

Despite some similarities with H2, haplogroup H13 provides a contrasting phylogeographic pattern. H13 dates to the end of the Last Glacial Maximum, ~19 [13.8–23.5] ka, and divides into three main clades, all dating to the early Late Glacial period: H13a (~17 [10.3–24.1] ka), H13b (~17 [11.5–21.9] ka) and H13c (~16 [10.7–21.6] ka). Although, once again, the majority of modern sequences are European, there is a considerable fraction of samples from more easterly regions (namely, the Caucasus, the Near East and South Asia), accounting for ~15% of our modern dataset (20% excluding sequences of unknown geographic origin) – more than twice as many as is the case for H2. The oldest H13 mitogenomes known are from the early Holocene of Georgia (Jones et al. 2015) and the famous Iron Gates Mesolithic in Romania and Serbia, close to 10 ka (Mathieson et al. 2018). It reappears ~5 ka as H13a1a in Yamnaya and Poltavka pastoralist burials from Russia (Haak et al. 2015; Mathieson et al. 2015), in Bell Beaker burials from Germany ~4 ka (Brotherton et al. 2013; Haak et al. 2015), and a few hundred years later in the Minoan Bronze Age of Crete (Lazaridis et al. 2017).

Interestingly, as with H2, we also found a possible signal of Bronze Age arrivals to the Indian Subcontinent in H13. H13a1a (~7 [5.6–8.8] ka) may have arisen in the Caucasus, but is largely a typical North European branch, with Middle Bronze Age samples from the Poltavka Culture (the successor to the Yamnaya in the Samara region in Russia, I0374: 2800–2200 BCE) (Mathieson et al. 2015) and England (I7572: 1510–1302 cal. BCE) (Olalde et al. 2018) at its root, and includes a Russian Yamnaya sequence in a deep Eastern European subclade (Haak et al. 2015). In our dataset, one modern Indian sample (Palanichamy et al. 2004) clusters within H13a1a, together with a Sintashta sequence (I0980: 2050–1650 BCE) from Russia (Narasimhan et al. 2018), far to the east to the north of the Aral Sea (H13a1a+12771) (Figure 2a). Thus H13a1a (similarly to H2b) likely participated in the long-range dispersals associated with the Yamnaya expansions, both westwards across northern Europe and eastwards into Central Asia, showing that, despite the significant male bias in the composition of these dispersals, some women at least were involved. H13a1a in India is reportedly found predominantly amongst the present-day Indo-Aryan-speaking Uttar Pradesh Brahmin communities (Palanichamy et al. 2015). We focus here, however, on H13a2a, which arose ~12 [8.1–15.5] ka, probably somewhere between

the South Caucasus and Iran, and is very unusual amongst haplogroup H sub-clusters in harbouring a diverse range of South Asian lineages.

The earliest archaeological evidence of the Neolithic in the Indian Subcontinent is from the site of Mehrgarh, dating to before 9 ka in Baluchistan (Jarrige and Jarrige 2006; Petrie 2015), to the west of the Indus Valley. This region was more moist in the early Holocene than it is today, with humidity values peaking ~8 ka (Van Campo 1986), providing good conditions for agriculture and allowing for the eventual rise of the Harappan Civilisation (Coningham and Young 2015). The analysis of non-autochthonous South Asian mtDNA haplogroups (i.e. lineages introduced towards the end of the Pleistocene and during the Holocene, as opposed to those that arose indigenously within the descendants of the earliest *Homo sapiens* settlers who dispersed from Africa around 60 ka) corroborates the archaeological record, with several lineages, including H13a2a, evidently entering the Subcontinent in this period from a putative source in the eastern end of the Fertile Crescent, potentially associated with the Dravidian language family (Silva et al. 2017).

H13a2a1 (dating to ~10 [5.3–13.7] ka) is shared between Iran and the Subcontinent and a previously unnamed H13a2a sub-cluster, labelled here as H13a2a2, is a Sri Lanka (Tamil, i.e. Dravidian-speaking) branch (The 1000 Genomes Project Consortium 2015) dating to ~7 ka [2.1–12.4] (albeit with wide confidence intervals, as it is a small sub-cluster). Moreover, H13a2a harbours two recently released ancient lineages (Narasimhan et al. 2018) that corroborate this scenario: one from Early Neolithic Turkmenistan, from the site of Tepe Anau (I4085) and dating to ~4000–3000 BCE (well before the pastoralist expansions from the north), and another from Iron Age Pakistan (I7717: 350–400 BCE) (Narasimhan et al. 2018) (Figure 2b). Therefore, H13 arrived in South Asia initially from Iran with agricultural dispersals, in the form of H13a2a, where it is still present today. It entered again later in the Bronze Age, when a different branch (H13a1a) was picked up during the Indo-European movements from the Steppe.

In a context of male-biased population movements, such as in the case of the Late Bronze Age movements into South Asia (Silva et al. 2017), evidence in the maternal gene pool is expected to be scarce. Nevertheless, we show here how a detailed phylogeographic approach, combining both modern and ancient variation, can provide additional clues into population movements, even in the case of strong male bias.

We thank Vagheesh Narasimhan, Swapan Mallick and David Reich for kindly providing *bam* files for the ancient mtDNA samples reported in Narasimhan et al. (2018). M.S., P.J., S.R., G.O.-G., K.D., G.F., A.F., B.Y., M.P., and M.B.R. received support from the Leverhulme Trust Doctoral Scholarship programme. This work

was partially supported by the Portuguese Foundation for Science and Technology (FCT), through the project PTDC/EPH-ARQ/4164/2014 partially funded by European Regional Development Fund (FEDER) (COMPETE 2020 project 016899). P.S. was supported by FCT, ESF and POPH through the FCT Investigator Programme (IF/01641/2013) and acknowledges FCT IP and ERDF (COMPETE2020 – POCI) for the CBMA strategic programme UID/BIA/04050/2013 (POCI-01-0145-FEDER-007569). T.R. is supported by an FCT grant (SFRH/BPD/108126/2015) and acknowledges the project [NORTE-01-0145-FEDER-000013], supported by NORTE 2020-Portugal 2020, through FEDER.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of Bronze Age Eurasia. Nature 522:167–172.

Anthony DW, Bogucki P, Comşa E, Gimbutas M, Jovanović B, Mallory JP, and Milisaukas S. 1986. The "Kurgan Culture," Indo-European origins, and the domestication of the horse: A reconsideration. Curr Anthropol. 27:291–313.

Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, Moller-Rieker S, Meller H, Ganslmeier R, Friederich S, et al. 2013. Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. Science (80). 342:257–261.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Adler CJ, Richards SM, Der Sarkissian C, Ganslmeier R, Friederich S, et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun 4:1764.

Coningham R, Young R. 2015. The Archaeology of South Asia: From the Indus to Asoka, c. 6500 BCE–200 CE. New York (NY): Cambridge University Press.

Derenko M, Malyarchuk B, Denisova G, Perkova M, Litvinov A, Grzybowski T, Dambueva I, Skonieczna K, Rogalla U, Tsybovsky I et al. 2014. Western Eurasian ancestry in modern Siberians based on mitogenomic data. BMC Evol Biol 14:217.

Gimbutas M. 1963. The Indo-Europeans: Archeological problems. Am Anthropol. 65:815–836.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211.

Jarrige J-F, Jarrige C. 2006. Premiers pasteurs et agriculteurs dans le sous-continent Indo-Pakistanais. Comptes Rendus Palevol 5:463–472.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. 2015 Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun 6:8912.

Juras A, Chyleński M, Ehler E, Malmström H, Żurkiewicz D, Włodarczak P, Wilk S, Peška J, Fojtík P, Králík M, et al. 2018 Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. Sci Rep 8:11603.

Krzewińska M, Merve Kılınç G, Juras A, Koptekin D, Chyleński M, Nikitin AG, Shcherbakov N, Shuteleva I, Leonova T, Kraeva L, et al 2018 Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. Sci Adv 4: eaat4457.

Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, Furtwängler A, Peltzer A, Posth C, Vasilakis A, et al. 2017. Genetic origins of the Minoans and Mycenaeans. Nature 548:214-218.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. 2016 Genomic insights into the origin of farming in the ancient Near East. Nature 536:419–424.

Li S, Besenbacher S, Li Y, Kristiansen K, Grarup N, Albrechtsen A, Sparsø T, Korneliussen T, Hansen T, Wang J, et al. 2014. Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. Eur J Hum Genet 22:1040–1045.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5:13.

Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H-V, Parik J, et al. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol. 21:2012–2021.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503.

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Candilio F, Cheronet O, et al. 2018. The genomic history of southeastern Europe. Nature. 555:197–203

Narasimhan VM, Patterson NJ, Moorjani P, Lazaridis I, Mark L, Mallick S, Rohland N, Bernardos R, Kim AM, Nakatsuka N, et al. 2018. The Genomic Formation of South and Central Asia. bioRxiv 292581.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szécsényi-Nagy A, Mittnik A, et al. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature 555:190–196.

Pala M, Soares P, Richards MB 2016. Archaeogenetic and palaeogenetic evidence for metal age mobility in Europe. In: Koch, JT Koch, Cunliffe B, editors. Celtic from the West. Vol. 3. Atlantic Europe in the Metal Ages – questions of shared language. Oxford: Oxbow Books; p321–384.

Palanichamy MG, Mitra B, Zhang C-L, Debnath M, Li G-M, Wang H-W, Agrawal S, Chaudhuri TK, Zhang Y-P 2015. West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. Hum. Genet. 134:637–647.

Palanichamy MG, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhuri TK, Palla V, Zhang YP. 2004 Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75:966–978.

Parpola A. 2015. The Roots of Hinduism: The Early Aryans and the Indus Civilization. New York (NY): Oxford University Press.

Petrie CA. 2015. Case Study: Mehrgarh. In The Cambridge World History. Volume 2. A World with Agriculture, 12,000 BCE–500 CE. In Barker CA, Goucher C, editors. Cambridge:Cambridge University Press, p. 289–309.

Silva M, Oliveira M, Vieira D, Brandao A, Rito T, Pereira JB, Fraser RM, Hudson B, Gandini F, Edwards C, et al. 2017. A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. BMC Evol Biol. 17:88.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards, M.B. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 84:740–759.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, Richards MB. 2010. The archaeogenetics of Europe. Curr Biol. 20:R174-83.

Staubwasser M, Sirocko F, Grootes PM, Segl M. 2003. Climate change at the 4.2 ka BP termination of the Indus valley civilization and Holocene south Asian monsoon variability. Geophys Res Lett. 30.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68–74.

van Campo E. 1986. Monsoon fluctuations in two 20,000-Yr B.P. oxygen-isotope/pollen records off Southwest India. Quat Res. 26:376–388.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:E386-94.

Wang C-C, Reinhold SR, Kalmykov A, Wissgott A, Brandt G, Jeong C, Cheronet O, Ferry M, Harney E, Keating D, et al. 2018. The genetic prehistory of the Greater Caucasus. bioRxiv 322347.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics. 13:555–556.

Figure 1. Schematic tree of mtDNA subclade H2b.

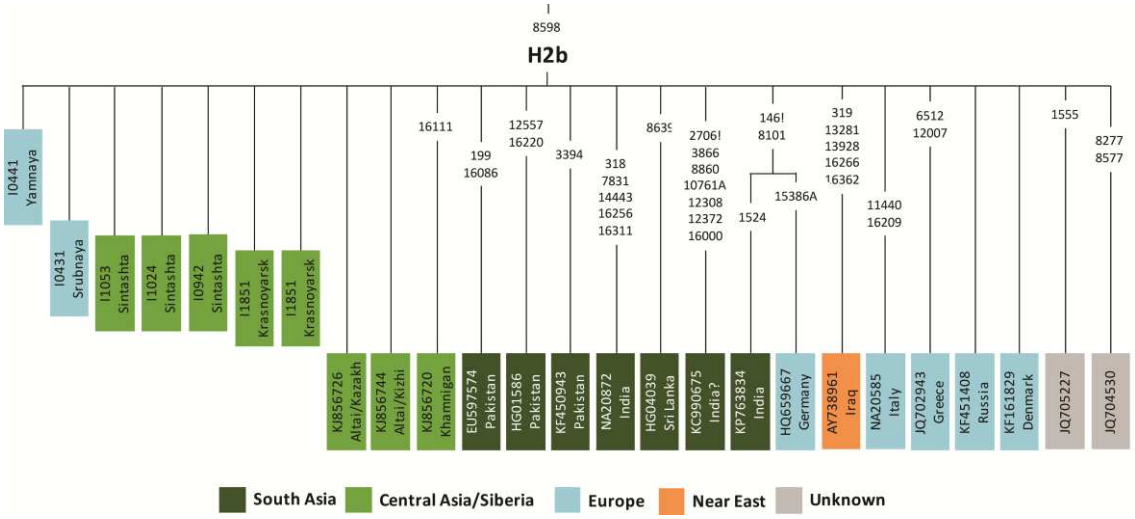

Figure 2. Schematic trees of mtDNA subclades H13a1a (only basal samples and H13a1a+12771 shown) (a) and H13a2a (b). Abbreviations for ancient samples: MBA – Middle Bronze Age, EN – Early Neolithic, IA – Iron Age.